LEVEL (12) B.S.

DEPARTMENT OF STATISTICS

The Ohio State University

OSU

COLUMBUS, OHIO

80 3 17 210

A MULTIVARIATE SOLUTION
OF THE MULTIVARIATE
RANKING AND SELECTION PROBLEM

by
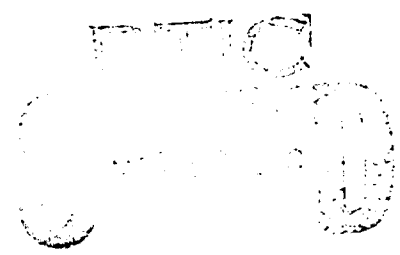
Edward J. Dudewicz*

and

Vidya S. Taneja

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 163a	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A MULTIVARIATE SOLUTION OF THE MULTIVARIATE RANKING AND SELECTION PROBLEM		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Edward J. Dudewicz Vidya S. Taneja		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0543
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics The Ohio State University Columbus, Ohio 43210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-403
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia 22217		12. REPORT DATE February 1980
		13. NUMBER OF PAGES i + 20
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approval for public release; distribution unlimited.

TR-163A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Key Words and Phrases: Ranking and selection; multivariate analysis; Heteroscedastic Method; multiple criteria; multiple objective; decision theory; Mahalanobis distance; multiple correlation; generalized variance; multidimensional utility; utility functions.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of selection of the best multivariate population is given a new formulation which does not involve reducing the populations to univariate quantities. This formulation's solution is developed for known, and (using the Heteroscedastic Method) also for unknown, variance-covariance matrices. Preference reversals and arbitrary nonlinear preference functions are explictly allowed in this new theory.

DD FORM 1473
1 JAN 73

A MULTIVARIATE SOLUTION OF THE
MULTIVARIATE RANKING AND SELECTION PROBLEM

Edward J. Dudewicz   and   Vidya S. Taneja

Department of Statistics   Department of Mathematics
The Ohio State University   Western Illinois University
Columbus, Ohio 43210   Macomb, Illinois 61455

ABSTRACT

The problem of selection of the best multivariate population
is given a new formulation which does not involve reducing the
populations to univariate quantities. This formulation's
solution is developed for known, and (using the Heteroscedastic
Method) also for unknown, variance-covariance matrices. Prefer-
ence reversals and arbitrary nonlinear preference functions are
explicitly allowed in this new theory.

1.   INTRODUCTION

Let π_i be a multivariate normal population with p (≥ 1)
variates, mean vector μ_i, and positive definite variance-
covariance matrix Σ_i (i = 1,...,k); i.e., let π_i be N_p (μ_i, Σ_i)
(i = 1,...,k). The problem of selecting that one of $\pi_1,...,\pi_k$
which is "best" in some precise sense is an important one which
arises frequently in practice (e.g. see Chapter 15, pp. 341-394,

1

of Gibbons, Olkin, and Sobel (1977)). However, "The whole field is as yet undeveloped and the reader [of Gibbons, Olkin, and Sobel (1977), p. 390] is encouraged to regard this chapter [their Chapter 15] as an introduction to a wide area that will see considerable development in the future as more meaningful models are formulated." This current paper deals with the development of such a model.

Most often, studies in this area have started by assuming given a known function $\phi = \phi(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ which is real-valued. One then defines $\phi_1 = \phi(\underset{\sim}{\mu}_1, \underset{\sim}{\Sigma}_1), \ldots, \phi_k = \phi(\underset{\sim}{\mu}_k, \underset{\sim}{\Sigma}_k)$, defines the population associated with the largest [sometimes, the smallest] of $\phi_1, \ldots, \phi_k$ as "best", and then proceeds to develop ranking and selection procedures for the resulting univariate problem. For example, one often then seeks to develop a procedure $\mathcal{G}$ such that, for a fixed P^* $(1/k < P^* < 1)$ the probability of a correct selection satisfies

$$\inf_{\Omega_p} P(CS | \mathcal{G}) = P^*$$

where Ω_p is some subset of the parameter space Ω of $\underset{\sim}{\mu}_1, \ldots, \underset{\sim}{\mu}_k$, $\underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_k$. (Let $\phi_{[1]} \leq \cdots \leq \phi_{[k]}$ denote $\phi_1, \ldots, \phi_k$ in numerical order. Then if "CS" means selection of the population with $\phi_{[k]}$ and Ω_p is a proper subset of Ω, we call Ω_p the preference zone and say we are dealing with the indifference-zone setting pioneered by Bechhofer (1954). While if "CS" means selection of a subset S of $\{\pi_1, \ldots, \pi_k\}$ such that S contains the population with $\phi_{[k]}$ and $\Omega_p = \Omega$, we say we are dealing with the subset-selection setting pioneered by Gupta (1956).) Various authors have made various choices of the real-valued function $\phi(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$. For example, $\phi = \underset{\sim}{\mu}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{\mu}$ (Alam and Rizvi (1965), Gupta (1966), Srivastava and Taneja (1972)), $\phi = \underset{\sim}{\mu}' \underset{\sim}{\mu}$ (Srivastava and Taneja (1972)), $\phi = \underset{\sim}{c}' \underset{\sim}{\mu}$ for a vector of constants $\underset{\sim}{c}$ (Krishnaiah and Rizvi (1966)), the generalized variance (Gnanadesikan and Gupta (1970), Regier (1976)), the multiple

2

correlation coefficient (Gupta and Panchapakesan (1969), Rizvi and Solomon (1973), Alam, Rizvi and Solomon (1976)), the product moment correlation for the bivariate (p = 2) case (Govindarajulu and Gore (1971)), and correlated variances (Arvesen and McCabe (1973)).

While in studies to date, as detailed above, the multi-variate ranking and selection problem has been univariatized (often for mathematical convenience rather than for compelling reasons why one should regard the $\phi(\mu, \Sigma)$ used as a natural measure of population goodness; while we do not mean to say or imply that the ϕ considered in the literature are always un-reasonable in practice, we do believe there has been somewhat more emphasis on them than their applicability warrants, due largely to their mathematical tractability; for example, we are not aware of a development of Mahalanobis distance showing it to be a natural measure of distance), in the present paper we attempt to give a first multivariate solution of the multi-variate ranking and selection problem. In particular, our treatment allows for such occurrences as (letting ">" denote "is better than" or "is preferred to")

$$\pi_1 \succ \pi_2 \succ \pi_3 \succ \pi_1 \tag{1.1}$$

which are well-known in (e.g.) sports (tennis, golf, etc.). While (1.1) is an anomaly in previous treatments of multivariate ranking and selection, it is to be expected in truly multi-variate problems (i.e., problems in which one cannot associate a univariate measure of goodness, or number, $\phi(\mu, \Sigma)$, with a given population, but must rather compare different (μ, Σ) pairs them-selves in order to determine which is preferred), and it is not anomalous in our treatment of multivariate ranking and selection. For other approaches to "anomalies" such as (1.1) (besides the most common one, of denying its existence and blaming it on "random variability") see Good and Tideman (1976), Fishburn (1977b), Young and Levenglick (1978), and Lee and Dudewicz (1980).

3

Note, however, that the approach presented below is also applicable to the situations (perhaps more common than those described in the above paragraph) where there is associated with each population π_i a numerical "goodness" $u_i = u(\mu_i)$, $1 \leq i \leq k$, such that π_j is preferred to π_ℓ iff $u_j > u_\ell$ ($1 \leq j$, $\ell \leq k$). In that situation, we are able to allow not simply linear functions $u(\mu_i)$, but arbitrarily complex reality, including e.g. quadratic, polynomial, exponential, and power series or even Fourier series functions.

Due to repeated skepticism of the true existence of situations like (1.1), and as a contrast with the situation in the above paragraph, an example of (1.1) may be useful at this point. Namely, in the Open Tennis Career Records as of 11/14/78 (compiled by Mr. Steve Flink of World Tennis, and graciously provided by Mr. Ron Bookman) one can let π_1 = Martina Navratilova, π_2 = Virginia Ruzici, and π_3 = Regina Marsikova. Then $\pi_1 \succ \pi_2$ (by 3-0), $\pi_2 \succ \pi_3$ (by 4-1), but $\pi_3 \succ \pi_1$ (by 2-1). Any reasonable approach assigning $u(\mu_i)$ independent of examination of μ_j's would, in this setting, come to erroneous conclusions as to the relative merits of π_1 and π_3.

Briefly, in our new theory we let $g(\mu_1,\ldots,\mu_k)$ be an experimenter-specified function with range space $\{1, 2,\ldots,k\}$ and such that

$$g(\mu_1,\ldots,\mu_k) = j \tag{1.2}$$

iff, given a choice among $\mu_1,\ldots,\mu_k$, the experimenter would prefer μ_j. Below we give multivariate preference selection procedures $\Psi_{M'p}$ which sample from each of $\pi_1,\ldots,\pi_k$, estimate μ_i by $\hat{\mu}_i$ ($i = 1,\ldots,k$), and select

$$\pi_g(\hat{\mu}_1,\ldots,\hat{\mu}_k). \tag{1.3}$$

This is done first for the problem of selecting π_b where $b = g(\mu_1,\ldots,\mu_k)$ when $\Sigma_1 = \ldots = \Sigma_k = \sigma^2 I$ with σ^2 known (Section 2), then when $\Sigma_1 = \ldots = \Sigma_k = \Sigma$ with Σ known (Section 3), next for

4

the case of $\mathfrak{f}_1,\ldots,\mathfrak{f}_k$ known but unequal (Section 4), and finally
(using the previous results as a base for The Heteroscedastic
Method) in the general case when $\mathfrak{f}_1,\ldots,\mathfrak{f}_k$ are unknown and not
necessarily equal (Section 5). (While sections 2, 3, and 4
could be combined, we have kept their results separated for
several reasons. First, our new approach to multivariate selec-
tion is most easily understood in the context of Section 2,
where the fewest multivariate complexities are present. Second,
the setting of Section 3 with $\mathfrak{f} = \ldots = \mathfrak{f}$ is that which The Hetero-
scedastic Method builds its solution on. The probability
expressions developed in Section 4 are useful in Section 5.)
Extensions are then noted (Section 6).

We wish to emphasize that the nature of the preference
function g is not a chief point of interest in this paper;
however, our work holds for any and every such function. The
specification and/or elicitation of such functions is part of
the field of decision theory. Up to now it has dealt with only
relatively "simple" types of functions. (For some illustrative
papers developing nonadditive (but still relatively simple)
utility theory, see Farquhar (1976) and Fishburn (1977a). For
some typical applications of this utility theory in practical
situations, see Giauque and Peebles (1976) and Krischer (1976).
For some consideration of assessment of multiattribute cardinal
utility functions see Kirkwood (1976). For research papers and
a survey of the field of multiple criteria/objective decision
making, see respectively Starr and Zeleny (1977) and Hwang and
Masud (1979).) We plan to investigate such specifications as
(e.g.) polynomial functions g. The present paper investigates
how one can do statistical inference with any such function
(even if not completely elicited or known), and this has been
hitherto unavailable.

5

2. SELECTION OF THE BEST WHEN $\mathfrak{t}_1 = \ldots = \mathfrak{t}_k = \sigma^2 I$, σ^2 KNOWN

Let π_i be $N_p(\mu_i, \mathfrak{t}_i)$ for $i = 1, \ldots, k$, and assume $p \geq 1$ and $\mathfrak{t}_1 = \ldots = \mathfrak{t}_k = \sigma^2 I$. Here I is the $p \times p$ identity matrix and and σ^2 is assumed known. Let $g(\mu_1, \ldots, \mu_k)$ be an experimenter-specified function with range space $\{1, 2, \ldots, k\}$, where (1.2) denotes the fact that among $\mu_1, \ldots, \mu_k$ the experimenter prefers μ_j. The natural selection procedure in this setting is to observe n independent observations from each of $\pi_1, \ldots, \pi_k$, estimate μ_i by the usual sample mean vector $\hat{\mu}_i$ ($1 \leq i \leq k$), and select

$$\pi_{g(\hat{\mu}_1, \ldots, \hat{\mu}_k)} \, . \tag{2.1}$$

Call this <u>selection procedure</u> $\mathcal{C}_{MVP}(\sigma^2 I)$.

Of course if the true means

$$\mu = (\mu_1, \ldots, \mu_k) \tag{2.2}$$

are such that $g(\mu) = j$ while $g(\mu + \varepsilon) = \ell$ ($\ell \neq j$) for a matrix ε of small numbers, then the probability of correct selection of $\mathcal{C}_{MVP}(\sigma^2 I)$,

$$P(CS | \mathcal{C}_{MVP}(\sigma^2 I)) \, , \tag{2.3}$$

will not be much larger than $1/k$. We will therefore show next how to specify the sample size n per population so that, for a reasonable preference zone $\Omega_P(d^*)$ ($d^* > 0$) and a fixed P^* ($1/k < P^* < 1$) procedure $\mathcal{C}_{MVP}(\sigma^2 I)$ satisfies

$$\inf_{\Omega_P(d^*)} P(CS | \mathcal{C}_{MVP}(\sigma^2 I)) \geq P^* \, . \tag{2.4}$$

Let

$$P_j = \{\mu : \ g(\mu) = j\} \, , \quad j = 1, \ldots, k, \tag{2.5}$$

and note that $P_1, \ldots, P_k$ are disjoint preference sets whose union is $\mathbb{R}^{kp}$. Define the distance between any two points a and b of $\mathbb{R}^{kp}$ as the usual Euclidean distance

6

$$d(\underset{\sim}{a},\ \underset{\sim}{b}) = (\ \sum_{1}^{kp} (a_i - b_i)^2)^{1/2}\ , \qquad\qquad (2.6)$$

and denote the distance from $\underset{\sim}{\mu}$ to the boundary of $P_{g(\underset{\sim}{\mu})}$ by

$$d_B(\underset{\sim}{\mu}) = \inf\{d(\underset{\sim}{\mu},\ \underset{\sim}{b}): \quad \underset{\sim}{b} \notin P_{g(\underset{\sim}{\mu})}\}\ . \qquad\qquad (2.8)$$

We now set our <u>probability requirement</u> for any procedure ϑ as

$$P(CS|\mathcal{G}) \geq P^* \text{ whenever } d_B(\underset{\sim}{\mu}) \geq d^*\ .$$

(That is, we desire a selection procedure $\mathcal{G}$ which has probability at least P^* of choosing the true best population (event "CS") whenever the mean vector $\underset{\sim}{\mu}$ is at least distance d^* from mean

vectors where other populations are best. This same require- ment is used in Sections 3 and 4 below also.) Whenever $d_B(\underset{\sim}{\mu}) \geq d^*$ we have

$$P(CS|\theta_{MVP}(\sigma^2 I)) = P(\hat{\underset{\sim}{\mu}}\ \varepsilon\ P_{g(\underset{\sim}{\mu})})$$

$$\geq P[d(\underset{\sim}{\mu},\ \hat{\underset{\sim}{\mu}}) \leq d_B(\underset{\sim}{\mu})]$$

$$\geq P[d(\underset{\sim}{\mu},\ \hat{\underset{\sim}{\mu}}) \leq d^*] \qquad\qquad (2.9)$$

$$= P[\ \sum_{i=1}^{k}\ \sum_{\ell=1}^{p} (\hat{\mu}_{i\ell} - \mu_{i\ell})^2 \leq (d^*)^2]$$

$$= P[\ \sum_{i=1}^{k}\ \sum_{\ell=1}^{p} ((\hat{\mu}_{i\ell} - \mu_{i\ell})/(\sigma/\sqrt{n}))^2 \leq n(d^*)^2/\sigma^2]$$

$$= P[Y \leq n(d^*)^2/\sigma^2] \qquad\qquad (2.9a)$$

where Y has a (central) chi-square distribution with kp degrees of freedom. Hence we obtain

<u>THEOREM 1.</u> Selection procedure $\mathcal{G}_{MVP}(\sigma^2 I)$ satisfies probability requirement (2.8) if the sample size n per population satisfies

7

$$n \geq \chi^2_{kp} (P^*) \, \sigma^2/(d^*)^2 \qquad (2.10)$$

where $\chi^2_{kp}(P^*)$ is the value a central chi-square random variable with kp degrees of freedom fails to exceed with probability P^*.

The problem of how a practitioner will in practice choose P^* and δ^* deserves some comment. First, choice of P^* is similar to choice of power in tests of hypotheses, and a common desire is for $P^* = 0.95$ or some similarly high value. The choice of δ^* is similar to the specification in ANOVA of the alternative (noncentrality parameter) at which one wishes the power guarantee. If the experimenter specifies the minimum range between the largest of $\mu_{1,i}$, $\mu_{2,i}, \ldots, \mu_{k,i}$ and the smallest of $\mu_{1,i}$, $\mu_{2,i}, \ldots, \mu_{k,i}$ which he wishes to detect as $\Delta_i > 0$ units, then whenever

$$\max(\mu_{1,i}, \mu_{2,i}, \ldots, \mu_{k,i}) - \min(\mu_{1,i}, \mu_{2,i}, \ldots, \mu_{k,i}) \geq \Delta_i \quad (2.11)$$

we have (for any two possible μ's, say $\underset{\sim}{a}$ and $\underset{\sim}{b}$, which satisfy (2.11)) $d(\underset{\sim}{a}, \underset{\sim}{b}) \geq \Delta_i/2$. Hence (since the Δ_i of interest may vary from component to component), one may choose $\delta^* = \min(\Delta_1, \ldots, \Delta_p)/2$. This is similar to the ANOVA situation as considered by, e.g., Scheffé (1959), pp. 63-64.

Regarding use of the usual sample mean vector $\hat{\underset{\sim}{\mu}}_i$ to estimate $\underset{\sim}{\mu}_i$ ($1 \leq i \leq k$), it should be noted that this is perhaps not optimal. However, recent results of Draper and Van Nostrand (1977) suggest an optimal estimator will not be far away from this estimator.

Note that use of $d(\underset{\sim}{\mu}, \underset{\sim}{b})$ specifies the part of $\mathbb{R}^{kp}$ where we control the P(CS); it does not univariatize our preferences. Also, g as a function of the $\underset{\sim}{\mu}_i$'s alone is most usual when the $\underset{\sim}{\mu}_i$'s relate to long-term yield (in decision theory); extension to

8

functions of μ_i's and Σ_i's (when Σ_i's are unknown; when they are known, the present results suffice) is a desirable area for further study.

3. SELECTION OF THE BEST WHEN $\Sigma_1 = \ldots = \Sigma_k = \Sigma$, Σ KNOWN

Let π_i by $N_p (\mu_i, \Sigma_i)$ for $i = 1, \ldots, k$, and assume $p \geq 1$ and $\Sigma_1 = \ldots = \Sigma_k = \Sigma$. Here Σ is the common $p \times p$ variance-covariance matrix and is assumed positive-definite and known. Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$ denote the characteristic roots of Σ. Let $g(\mu_1, \ldots, \mu_k)$ be an experimenter-specified function with range space $\{1, \ldots, k\}$ as in Section 2, and consider using the same selection procedure as in Section 2 (but now denote it by $\mathcal{R}_{MVP}(\Sigma)$ to explicitly take notice of the different variance-covariance structure being assumed). Still taking our probability requirement as (2.8), we wish to determine the common sample size n per population in such a way that (2.8) is guaranteed. We will find the following result (Rao (1965), p. 50) useful.

LEMMA 1. Let A be any symmetric $p \times p$ matrix, let $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_p$ denote the characteristic roots of A, and let $\underset{\sim}{x}$ be any $p \times 1$ vector. Then

$$\theta_1 \underset{\sim}{x}' \underset{\sim}{x} \leq \underset{\sim}{x}' A \underset{\sim}{x} \leq \theta_p \underset{\sim}{x}' \underset{\sim}{x} . \tag{3.1}$$

Now, choosing $A = \Sigma^{-1}$ in Lemma 1 (so θ_1 is the smallest characteristic root of Σ^{-1}, i.e. $1/\lambda_p$) we are in a position to show that whenever $d_B(\underset{\sim}{\mu}) \geq d^*$ we have

9

$$P(CS| \mathcal{G}_{MVP}(\ddagger)) = P(\hat{\underset{\sim}{\mu}} \in P_{g(\underset{\sim}{\mu})})$$

$$\geq P[\sum_{i=1}^{k} \sum_{\ell=1}^{p} (\hat{\mu}_{i\ell} - \mu_{i\ell})^2 \leq (d^*)^2]$$

$$= P[\sum_{i=1}^{k} (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i)' (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i) \leq (d^*)^2]$$

$$= P[\sum_{i=1}^{k} n \frac{1}{\lambda_p} (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i)' (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i) \leq \frac{n}{\lambda_p} (d^*)^2]$$

$$\geq P[\sum_{i=1}^{k} n(\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i)' \ddagger^{-1} (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i) \leq n(d^*)^2/\lambda_p]$$

$$= P[Y \leq n(d^*)^2/\lambda_p]$$

where Y has a (central) chi-square distribution with kp degrees of freedom. Hence we obtain

THEOREM 2. Selection procedure $\mathcal{G}_{MVP}(\ddagger)$ satisfies probability requirement (2.8) if the sample size n per population is such that

$$n \geq \chi^2_{kp}(P^*) \lambda_p/(d^*)^2 \tag{3.2}$$

where $\chi^2_{kp}(P^*)$ is the value a central chi-square random variable with kp degrees of freedom fails to exceed with probability P^* and λ_p is the largest characteristic root of $\ddagger$.

4. SELECTION OF THE BEST WHEN $\ddagger_1, \ldots, \ddagger_k$ KNOWN

Let π_i be $N_p(\underset{\sim}{\mu}_i, \ddagger_i)$ for $i = 1, \ldots, k$, and assume $p \geq 1$. Here the k positive-definite p×p variance-covariance matrices $\ddagger_1, \ldots, \ddagger_k$ are assumed known, but need not be equal. Let $g(\underset{\sim}{\mu}_1, \ldots, \underset{\sim}{\mu}_k)$ be an experimenter-specified function with range space $\{1, \ldots, k\}$ as in Section 2, and consider using the same selection procedure as in Section 2 (but now denote it by $\mathcal{G}_{MVP}(\ddagger_1, \ldots, \ddagger_k)$ to explicitly take notice of the different variance-covariance structures being assumed), with the

10

modification that now n_i observations are taken from π_i ($1 \le i \le k$). Still taking our probability requirement as (2.8), we wish to determine the sample sizes $n_1,\ldots,n_k$ in such a way that (2.8) is guaranteed.

Let λ_{ip} denote the maximum characteristic root of $\sharp_i$ (i.e., the smallest characteristic root of $\sharp_i^{-1}$), $i = 1,\ldots,k$. Using results developed previously, it then follows that whenever $d_B(\underset{\sim}{\mu}) \ge d^*$ we have (letting $n_{[1]} = \min(n_1,\ldots,n_k)$ and $\lambda_{[k]p} = \max(\lambda_{1p},\ldots,\lambda_{kp})$)

$P(CS | \mathscr{P}_{MVP}(\sharp_1,\ldots,\sharp_k))$

$$\ge P[\sum_{i=1}^{k} (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i)'(\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i) \le (d^*)^2]$$

(4.1)

$$\ge P[\sum_{i=1}^{k} n_i \frac{1}{\lambda_{ip}} (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i)' (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i) \le \frac{n_{[1]}}{\lambda_{[k]p}} (d^*)^2]$$

$$\ge P[\sum_{i=1}^{k} n_i (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i)' \sharp_i^{-1} (\hat{\underset{\sim}{\mu}}_i - \underset{\sim}{\mu}_i) \le n_{[1]}(d^*)^2/\lambda_{[k]p}]$$

$$= P[\sum_{i=1}^{k} Y_i \le n_{[1]}(d^*)^2/\lambda_{[k]p}]$$

where $Y_1,\ldots,Y_k$ are independent (central) chi-square random variables with p degrees of freedom. Hence we obtain

THEOREM 3. Selection procedure $\mathscr{P}_{MVP}(\sharp_1,\ldots,\sharp_k)$ satisfies probability requirement (2.8) if the sample sizes $n_1,\ldots,n_k$ are such that

$$n_{[1]} \ge \chi^2_{kp}(P^*) \lambda_{[k]p}/(d^*)^2 \tag{4.2}$$

where $\chi^2_{kp}(P^*)$ is the value a central chi-square random variable with kp degrees of freedom fails to exceed with probability P^*, $n_{[1]}$ is the smallest of $n_1,\ldots,n_k$, and $\lambda_{[k]p}$ is the largest of the characteristic roots of $\sharp_1,\ldots,\sharp_k$.

11

Using Theorem 3, in the design problem one would of course choose equal sample sizes $n_1 = n_2 = \ldots = n_k$ equal to the smallest integer $\geq$ the right-hand side of (4.2). However, if unequal sample sizes $n_1, \ldots, n_k$ have already been taken, one can nevertheless easily assess the smallest d^* for which (2.8) is satisfied by solving (4.2), as an equality, for d^*.

As an exact expression for the first bounding expression in (4.1), we have

$$P(CS | \mathcal{G}_{MVP}(\Sigma_1, \ldots, \Sigma_k)) \geq P[\sum_{i=1}^{k} (\hat{\mu}_i - \mu_i)' (\hat{\mu}_i - \mu_i) \leq (d^*)^2]$$

$$\tag{4.3}$$

$$= P[S_1 + \ldots + S_k \leq (d^*)^2]$$

where $S_1, \ldots, S_k$ are independent random variables. This can be used to achieve (2.8) with a smaller total sample size $n_1 + n_2 + \ldots + n_k$ in the design problem, via choosing $n_1, \ldots, n_k$ so that (4.3)'s right-hand side equals P^* and $n_1 + n_2 + \ldots + n_k$ is minimized. Similarly, in the situation where $n_1, \ldots, n_k$ have already been taken, the smallest d^* for which (4.3) equals P^* (hence (2.8) is satisfied) will be smaller than that provided by Theorem 3, hence yielding an improved (strengthened) probability requirement. However, since it is computationally difficult (i.e., expensive and time consuming) to evaluate (4.3), we have emphasized a simpler result (albeit it requires a larger sample size (in the design setting) and yields a larger d^* for (2.8) (in the data analysis setting)) in Theorem 3.

Since (4.3) is of interest in Section 5, we note that it may be calculated using the following facts. Since we may represent S_j as $S_j = Y_j' Y_j$ where $Y_j \sim N_p(0, \Sigma_j/n_j)$ $(1 \leq j \leq k)$, it follows that

12

$$H_j(t) \equiv P[S_j \leq t] = (2\pi)^{-p/2} |\sharp_j/n_j|^{-1/2} \int \ldots \int_{\underset{\sim}{y'_j}\underset{\sim}{y_j} \leq t} e^{-\frac{1}{2}\underset{\sim}{y'_j}(\sharp_j/n_j)^{-1}\underset{\sim}{y_j}} d\underset{\sim}{y_j}$$

(4.4)

$$= (2\pi)^{-p/2} \int \ldots \int_{\underset{r=1}{\overset{p}{\sum}} \lambda_{jr} w_r^2/n_j \leq t} e^{-\frac{1}{2}\underset{\sim}{w'}\underset{\sim}{w}} d\underset{\sim}{w}$$

(where we utilized an orthogonal transformation), and (from Pachares (1955)) that

$$H_j(t) = \frac{(n_j t/2)^{p/2}}{|\sharp_j/n_j|^{1/2}} \sum_{s=0}^{\infty} \frac{(-n_j t/2)^s \pi^{-p/2}}{\Gamma(\frac{1}{2}p + s + 1)} \left\{ \sum_{i_1 + \ldots + i_p = s} \frac{\Gamma(i_1 + \frac{1}{2}) \ldots \Gamma(i_p + \frac{1}{2})}{i_1! \ldots i_p! \lambda_{j_1}^{i_1} \ldots \lambda_{j_p}^{i_p}} \right\}.$$

(4.5)

5. SELECTION OF THE BEST WHEN $\sharp_1, \ldots, \sharp_k$ UNKNOWN, UNEQUAL

Let π_i be $N_p(\underset{\sim}{\mu_i}, \sharp_i)$ for $i = 1, \ldots, k$, and assume $p \geq 1$. Here the k positive-definite $p \times p$ variance-covariance matrices $\sharp_1, \ldots, \sharp_k$ are assumed unknown, and need not be equal. Let $g(\underset{\sim}{\mu_1}, \ldots, \underset{\sim}{\mu_k})$ be an experimenter-specified function with range space $\{1, \ldots, k\}$ as in Section 2, and consider the same selection problem as in Section 2. It follows from Dudewicz (1971) and Dudewicz and Dalal (1975) that no single-stage procedure θ for this problem can satisfy a probability requirement like (2.8), namely

$$P(CS|\theta) \geq P^* \text{ whenever } d_B(\underset{\sim}{\mu}) \geq d^*.$$

(5.1)

13

However, since (5.1) is the type of probability requirement one would usually desire in practice, it is desirable to seek a θ satisfying (5.1) in a broader class than the class of single-stage procedures.

Dudewicz and Bishop (1979) have given a method, called The Heteroscedastic Method, which allows us to modify procedure $\theta_{MVP}(\ddagger)$ of Section 3 into a procedure θ_{HM} which solves the present problem. Namely, let <u>procedure θ_{HM}</u> be specified by a sampling rule and a terminal decision rule as follows.

<u>Sampling Rule for θ_{HM}</u>. Select $z > 0$, and integer $n_0 > p$, and a $p \times p$ positive-definite matrix (α_{rs}). Take observations from each and every population π_c ($c = 1, \ldots, k$) as follows. Take n_0 initial observations $X_{c1}, \ldots, X_{cn_0}$ where $X_{ci} = (X_{c1i}, X_{c2i}, \ldots, X_{cpi}$ $(i = 1, 2, \ldots, n_0)$ and compute

$$\overline{X}_{ci} = \frac{1}{n_0} \sum_{\ell=1}^{n_0} X_{ci\ell}, \quad S_{cij} = \sum_{\ell=1}^{n_0} (X_{ci\ell} - \overline{X}_{ci})(X_{cj\ell} - \overline{X}_{cj}), \quad (5.2)$$

$$s_{cij} = \frac{1}{(n_0 - 1)} S_{cij}, \quad i, j = 1, 2, \ldots, p .$$

Define the positive integer N_c by

$$N_c = \max\{n_0 + p^2, [z^{-1} \sum_{i,j=1}^{p} \alpha_{ij} s_{cij}] + 1\}, \quad (5.3)$$

where $[q]$ denotes the largest integer less than q, and select p ($p \times N_c$) matrices

$$A_{cr} = \begin{bmatrix} a_{cr_{11}} & \cdots & a_{cr_{1N_c}} \\ \vdots & & \\ a_{cr_{p1}} & \cdots & a_{cr_{pN_c}} \end{bmatrix} \quad (r = 1, 2, \ldots, p)$$

in such a way that:

14

1) $a_{cr_{i1}} = \ldots = a_{cr_{in_0}}$;

2) $A_{cr} \eta_c = \varepsilon_r$ where η_c is the $N_c \times 1$ vector $(1,1,\ldots,1)'$ and ε_r is the $p \times 1$ vector whose $r\underline{th}$ element is 1 and all other elements are zero;

and

3) $A_c A_c' = z(\alpha^{rs}) \; \mathbf{c} \; (s_c^{ij})$, where $A_c' = (A_{c1}', A_{c2}', \ldots, A_{cp}')$, $\mathbf{c}$ denotes the direct product, and (b^{ij}) denotes the inverse of the matrix (b_{ij}), $r, i = 1, 2, \ldots, p$.

Next take $N_c - n_0$ additional observations $X_{c,n_0+1}, \ldots, X_{cN_c}$ and

compute

$$\tilde{X}_{cr} = \sum_{i=1}^{p} \sum_{\ell=1}^{N_c} a_{cr_{i\ell}} X_{ci\ell} \qquad (r=1,2,\ldots,p). \qquad (5.4)$$

For π_c construct the p-dimensional vector $\tilde{\bar{X}}_c = (\tilde{X}_{c1}, \ldots, \tilde{X}_{cp})$, $c = 1, 2, \ldots, k$.

Terminal Decision Rule for θ_{HM}. Take the same decision as rule θ_{MVP} ($\ddagger$) of Section 2 would when it took a sample size n per population, had $\ddagger/n = z(\alpha^{rs})$, and observed

$$(\hat{\mu}_1, \ldots, \hat{\mu}_k) = (\tilde{\bar{X}}_1, \ldots, \tilde{\bar{X}}_k) . \qquad (5.5)$$

That is, select

$$\pi_g(\tilde{\bar{X}}_1, \ldots, \tilde{\bar{X}}_k) . \qquad (5.6)$$

Now from (a slight extension of) Theorem (2.20) of Dudewicz and Bishop (1979) it follows that $\exists \; z > 0$ such that we have THEOREM 4. Selection procedure θ_{HM} satisfies probability requirement (5.1). The constant $z > 0$ in θ_{HM} is to be chosen so that

$$P[\; \sum_{i=1}^{k} (\tilde{\bar{X}}_i - \mu_i)' \; (\tilde{\bar{X}}_i - \mu_i) \leq (d^*)^2] = P^* . \qquad (5.7)$$

While the distribution of (5.5) is independent of $(\ddagger_1, \ldots, \ddagger_k)$, it is very complicated (see equations (2.12), (2.13) of Dudewicz and Bishop (1979)), hence calculation of

15

z > 0 which satisfies (5.7) is not a simple matter. However, for large n_0 (a design constant under the experimenter's control) we may approximate (5.7)'s solution:

THEOREM 5. As $n_0 \to \infty$ the $z > 0$ which solves (5.7) approaches the solution of (5.7) when $\tilde{X}_1, \ldots, \tilde{X}_k)$ is replaced by $(Y_1, \ldots, Y_k)$ where $Y_1, \ldots, Y_k$ are independent random variables and Y_i is $N_p(\mu_i, zp(\alpha^{rs}))$. This solution may be calculated from (4.3) and (4.5).

PROOF. This follows from the proof of Theorem (5.1) of Dudewicz and Bishop (1979), where we find that the limiting (as $n_0 \to \infty$) distribution of $(\tilde{X}_1, \ldots, \tilde{X}_k)$ is the same as the distribution of $(Y_1, \ldots, Y_k)$.

6. EXTENSIONS

While in this paper we have developed a new approach for one goal of the area known (see Gupta and Panchapakesan (1979), p. 7) as "ranking and selection", namely "selection of the best," problems such as the selection of the best t (of k) multivariate populations, as well as selection of a subset containing the best multivariate population, can also be given solutions using the methods of this paper. Details will appear elsewhere.

ACKNOWLEDGMENTS

referees, are acknowledged with thanks.

BIBLIOGRAPHY

Alam, K. and Rizvi, M. H. (1965). Selection from multivariate normal populations. Ann. Inst. Statist. Math., 18, 307-318.

Alam, K., Rizvi, M. H. and Solomon, H. (1976). Selection of the largest multiple correlation coefficient: Exact sample size case. Annals of Statistics, 4, 614-620.

Arvesen, J. N. and McCabe, G. P. (1975). Subset selection problems for variances with application to regression analysis. J. Amer. Statist. Assoc., 70, 166-170.

Bechhofer, R. E. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.

Chatterjee, S. K. (1959). On an extension of Stein's two-sample procedure to the multi-normal problem. Calcutta Statistical Association Bulletin, 8, 121-148.

Draper, N. R. and Van Nostrand, R. C. (1977). Ridge regression: Is it worthwile? Technical Report No. 501, Department of Statistics, The University of Wisconsin, Madison, Wisconsin.

Dudewicz, E. J. (1971). Nonexistence of a single-sample selection procedure whose P(CS) is independent of the variances. South African Statistical Journal, 5, 37-39.

Dudewicz, E. J. and Bishop, T. A. (1979). The heteroscedastic method. Optimizing Methods in Statistics (ed. J. S. Rustagi), 183-203, Academic Press, New York.

Dudewicz, E. J. and Dalal, S. R. (1975). Allocation of observations in ranking and selection with unknown variances. Sankhyā, 37B, 28-78.

Farquhar, P. H. (1976). Pyramid and semicube decompositions of multiattribute utility functions. Operations Research, 24, 256-271.

Fishburn, P. C. (1977a). Multicriteria choice functions based on binary relations. Operations Research, 25, 989-1012.

Fishburn, P. C. (1977b). Condorcet social choice functions. SIAM J. Appl. Math., 33, 469-489.

Giauque, W. C. and Peebles, T. C. (1976). Application of multidimensional utility theory in determining optimal test-treatment strategies for streptococcal sore throat and rheumatic fever. Operations Research, 24, 933-949.

Gibbons, J. D., Olkin, I. and Sobel, M. (1977). Selecting and Ordering Populations: A New Statistical Methodology, John Wiley & Sons, Inc., New York.

Gnanadesikan, M. and Gupta, S. S. (1970). A selection procedure for multivariate normal distributions in terms of the generalized variances. Technometrics, 12, 103-117.

Good, I. J. and Tideman, T. N. (1976). From individual to collective ordering through multidimensional attribute space. Proc. R. Soc. Lond. A, 347, 371-385.

Govindarajulu, Z. and Gore, A. P. (1971). Selection procedures with respect to measures of association. Statistical Decision Theory and Related Topics (ed. S. S. Gupta and J. Yackel), 313-345, Academic Press, New York.

Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo. Ser. No. 150, Inst. of Statist., Univ. of North Carolina, Chapel Hill, N. C.

Gupta, S. S. (1966). On some selection and ranking procedures for multivariate normal populations using distance functions. Multivariate Analysis (ed. P. R. Krishnaiah), 457-475, Academic Press, New York.

Gupta, S. S. and Panchapakesan, S. (1969). Some selection and ranking procedures for multivariate normal populations. Multivariate Analysis II (ed. P. R. Krishnaiah), 475-505, Academic Press, New York.

Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations, John Wiley & Sons, Inc., New York.

Gupta, S. S. and Studden, W. J. (1970). On some selection and
ranking procedures with applications to multivariate
populations. Essays in Probability and Statistics
(R. C. Bose et al. ed.)327-338, Univ. of North Carolina
Press, Chapel Hill, N. C.

Hwang, C.-L. and Masud, A. S. M. (1979). Multiple Objective
Decision Making - Methods and Applications, A State-
of-the-Art Survey, Springer-Verlag Berlin Heidelberg,
Germany.

Kirkwood, C. W. (1976). Parametrically dependent preferences
for multiattributed consequences. Operations Research,
24, 92-103.

Krischer, J. P. (1976). Utility structure of a medical
decision-making problem. Operations Research, 24,
951-971.

Krishnaiah, P. R. and Rizvi, M. H. (1966). Some procedures for
selection of multivariate normal populations better than
a control. Multivariate Analysis (ed. P. R. Krishnaiah)
447-490, Academic Press, New York.

Lee, Y. J. and Dudewicz, E. J. (1980). Awardee selection
procedures, with special reference to the Frank Wilcoxon
and Jack Youden prizes. Technometrics, 22, 121-124.

Pachares, J. (1955). Note on the distribution of a definite
quadratic form. Annals of Mathematical Statistics,
26, 128-131.

Ramberg, J. S. (1977). Selecting the best predictor variate.
Commun. Statist., A6 (11), 1133-1147.

Rao, C. R. (1965). Linear Statistical Inference and Its
Applications, John Wiley & Sons, Inc., New York.

Regier, M. H. (1976). Simplified selection procedures for
multivariate normal populations. Technometrics, 18,
483-489.

Rizvi, M. H. and Solomon, H. (1973). Selection of largest
multiple correlation coefficient: asymptotic case.
J. Amer. Statist. Assoc., 68, 184-188. Corrigendum,
69, 288.

Scheffé, H. (1959). The Analysis of Variance, John Wiley & Sons,
Inc., New York.

Srivastava, M. S. and Taneja, V. S. (1972). Some sequential procedures for ranking multivariate normal populations. Ann. Inst. Statist. Math., 24, 455-464.

Srivastava, M. S. and Taneja, V. S. (1974). Some sequential procedures for a multivariate slippage problem. Metron, 31, 1-10.

Starr, M. K. and Zeleny, M. (Editors) (1977). Multiple Criteria Decision Making, North-Holland Publishing Company, Amsterdam.

Young, H. P. and Levenglick, A. (1978). A consistent extension of Condorcet's election principle. SIAM J. Appl. Math., 35, 285-300.